U.S. PATENT APPLICATION

Inventor(s):   Ciaran MURPHY

Invention: METHOD FOR DETERMINING MASTER OR SLAVE MODE IN
STORAGE SERVER SUBNET

NIXON & VANDERHYE P.C.
ATTORNEYS AT LAW
1100 NORTH GLEBE ROAD
8TH FLOOR
ARLINGTON, VIRGINIA 22201-4714
(703) 816-4000
Facsimile (703) 816-4100

SPECIFICATION

Method for determining master or slave mode in storage server subnet

Field of Invention

This invention relates to networks and to network attached storage, with particular, but not exclusive, usefulness in local area networks

Background of the Invention

Networks generally consist of a number of devices such as workstations, personal computers, servers, hubs. routers, bridges and switches linked together by physical cable or wireless links   The devices on the network operate in accordance with a protocol to enable recognition of communicating devices and control of the data or traffic between them   Networks may take various forms such as a local area network (LAN) or a wide area network (WAN)

There is a constant demand for increasing levels of storage and for flexible or expandable storage   It is desirable for devices to be able to configure themselves automatically, e g on start up, or after a new deployment for their mode of operation without the intervention of a network manager

Summary of the Invention

The present invention is directed towards providing a Storage Area Network in which a plurality of storage servers one or more of which are dual-purpose or dual-mode, are interconnected in a sub-network that appears to the main network as a single storage area. and to a method of enabling a storage server to determine automatically if it is directly connected to the main network and is to function as a master server of the subnet or whether it is a slave server   A dual-mode storage is one that can operate as either a file server or as a block storage device   There may be instances in complex networks or in clusters where more than one server is a master or where a cluster operates collectively as a master

Accordingly the invention provides a storage area network comprising a plurality of storage servers, a first of which is directly connected to a main network and others of which are connected to the first storage server, the first storage server having an internal protocol server that inserts an identifier into the protocol package so that the protocol package differs from the protocol of the main network and the first storage server supplies the protocol to the other storage servers which recognise the identifiers

The invention also provides a method of establishing mode of operation for a storage server in a storage area network, the storage area network comprising a plurality of storage servers a first of which is directly connected (as herein defined) to a main network and others of which are connected to the first storage server, and said server for which the mode of operation is to be established being capable of operating in a master or slave modes, the method comprising starting up, examining an incoming event to determine if it corresponds to network protocol or contains a particular protocol modification, and when the event corresponds to network protocol without the particular modification proceeding with further operation in master mode and inserting the particular protocol modification into events destined to other storage servers, or when the incoming event contains the particular protocol modification, proceeding with further operation in slave mode

Within the context of the disclosure 'directly connected' means such that normal events or protocols are experienced There may be switches or other intervening devices, and the connection may take any form

Brief Description of the Drawings

The invention is now described by way of example with reference to the accompanying drawings in which

Figure 1 is a simplified schematic illustration of a local area network incorporating an embodiment of the invention

Figure 2 is an illustration showing an embodiment with a master cluster

Figure 3 is a diagram of DHCP message format

Figure 4 is a flow diagram of the startup procedure and master/slave recognition

Detailed Description of Preferred Embodiment

One of the benefits of networks is the ability to access data remotely from its storage location, and storage servers are therefore very significant components   There is constant pressure to increase and upgrade storage facilities and to link together storage servers   When servers are changed or switched off it is desirable for the servers to be able to establish automatically how they are connected

In the present invention a plurality of storage servers are interconnected in the form of a network or subnet that is seen by the network to which it is attached as a single storage area   A particularly suitable form of server for this purpose is one that is dual-mode and which can be interconnected to other similar servers via Ethernet links

Referring now to Figure 1 such a storage area network 1 is shown connected to a local area network (LAN) 2   Also connected to the LAN are other devices represented in the drawing by a PC client 3 and a DHCP Server 4   In real circumstances LANs are usually very much more complex, and include many more attached devices including switches and hubs   In many instances the networks are managed by a supervisor and some of the individual devices may also be managed

Within the storage area network (SAN) a first storage server 5 is connected via a switch 6 to a plurality of other storage servers 7, 8 and 9

In the preferred embodiment the storage servers are preferably all dual mode, identical and of a type that can be interconnected via Ethernet links although other forms of interconnection and different servers may be used as long as they can respond to the

protocol that is used. In the example now described dynamic host communication protocol (DHCP) is used, but the invention is not limited to this protocol

The storage servers have multiple Ethernet ports, but storage server 5 does not route any packets between its ports but requires an output address per port Therefore only storage server 5 is directly accessed from the LAN and all the other storage servers are accessed indirectly via the storage server 5

In operation storage server 5 operates in a "master" mode in which it has its internal DHCP server turned on and in which it has a block-level view of all the storage within the SAN and operates in file-level mode The other storage servers 7,8 and 9 operate in a "slave" mode in which they have their internal DHCP servers turned off and operate in block-level mode

As far as another device on the LAN is concerned, such as PC client 3, it sees the aggregated storage of the SAN as a single large storage volume When it sends file input/output requests (using any appropriate protocol such as NFS or CIFS) to the SAN these are received by the master storage server 5 which converts the file-level request into a block-level request and sends this to the appropriate slave (7, 8 or 9) which then performs the relevant block-level read or write to its internal storage The block-level protocol is usually SCSI over IP, also termed iSCSI IP-based traffic from the LAN (such as DHCP from the LAN) does not appear on the SAN due to the passage via the master 5, although as indicated later DHCP packets will appear on the SAN from the master server It is desirable for the storage servers in the SAN to be able to determine for themselves whether they are the master or a slave, and hence whether to operate in file-level or block-level mode The present invention modifies an aspect of the normal protocol on the LAN (for example the DHCP protocol or alternatively BOOTP or any other), so that it appears differently when seen on the SAN For the purpose of illustration, a mechanism for this is explained further with reference to DHCP protocol, and Figures 3 and 4

Figure 3 shows the DHCP message format, in which the 'OPTIONS' portion can contain a variable message (The DHCP OPTIONS field in fact has the same format as the

VENDOR SPECIFIC AREA of a BOOTP message, so that area can be utilised in BOOTP protocol similarly to the following explanation for DHCP) The OPTIONS field may contain a plurality of entries Each item contains a one-octet type an optional one-octet length, and a multi octet value One type value (43 decimal) is defined as the Vendor-Specific option This may be used for transferring vendor-specific information between the DHCP client and DHCP server in either direction Within an implementation of the invention utilising DHCP this type value may be used to identify that the DHCP packet has been generated by a DHCP server within a storage area network It will be appreciated that other modifiable parts of the protocol or other events may be used in the procedure taught in this invention

An automated master/slave recognition procedure utilising this modified protocol is now described with reference to Figure 4

On power up all the storage servers in the SAN are configured to power up in slave mode with their internal DHCP servers turned off, as indicated by box 10

When a new device is connected to an existing network, or when it is started up after being down, it sends out a request to the network for an IP address There are various schemes by which addresses are allocated, one of which is via DHCP, another of which is via a discovery mechanism, for example a process known as ICMP Echo Request and Reply Having been turned on, the storage servers will all request IP addresses, as shown by box 11

Box 12 represents the step of receiving the IP address, which is followed by the step of the server checking the DHCP options field, or such other field and protocol or event as has been modified

The storage server that is directly connected to the LAN has its IP address assigned via the LAN and detects this by absence of the vendor-specific option in a DHCP lease message, or by detecting manual settings of the IP address Therefore it recognises it must be the master, and this corresponds to the YES answer to box 13

Having determined that it is the master, the storage server connected to the LAN turns on its own internal DHCP server (box 14), and also assigns a fixed IP address to its Ethernet port on the storage subnet. The internal DHCP server of the master then sends out packets on the storage subnet containing the modified protocol

The other storage servers which are still operating in slave mode then obtain their IP addresses on the storage subnet (SAN) from the DHCP server in the master. Thus at the box 13 step, these servers detect the special signal in the DHCP packet, a NO response to box 13, and remain in slave mode

Although the embodiment described is in the context of a LAN using DHCP and iSCSI, other forms of network may utilise the SAN described and its operation. Likewise other protocols or events may be used and modified in a corresponding way. In configurations where more than one storage server is to operate in master mode, the other master server will also detect that they are directly connected to the network. In some instances a plurality of servers may operate together in a master cluster to provide resilience so that in the event of a failure of one of them another takes over

With a cluster, all the servers in the cluster are connected to the main network and so experience the event or protocol on that network. The master cluster can be regarded as parallel connections between the LAN 2 and switch 6 of Figure 1 and will run cluster software. Figure 2 shows an arrangement with two servers in a master cluster. They may be interconnected so as to provide their cooperative or resilient functions. The interconnection protocol or cluster software may include only allowing one 'main' master turning on its DHCP server to send out the modified protocol, and for other servers of the cluster to run in file mode but only to turn on their DHCP servers in the event of a failure such that they are to become the new main master

It will be appreciated that start up in file level mode is also possible with the adoption of master mode corresponding to turning on the internal DHCP server. Detecting the modified protocol would then induce a switch to slave or block level mode